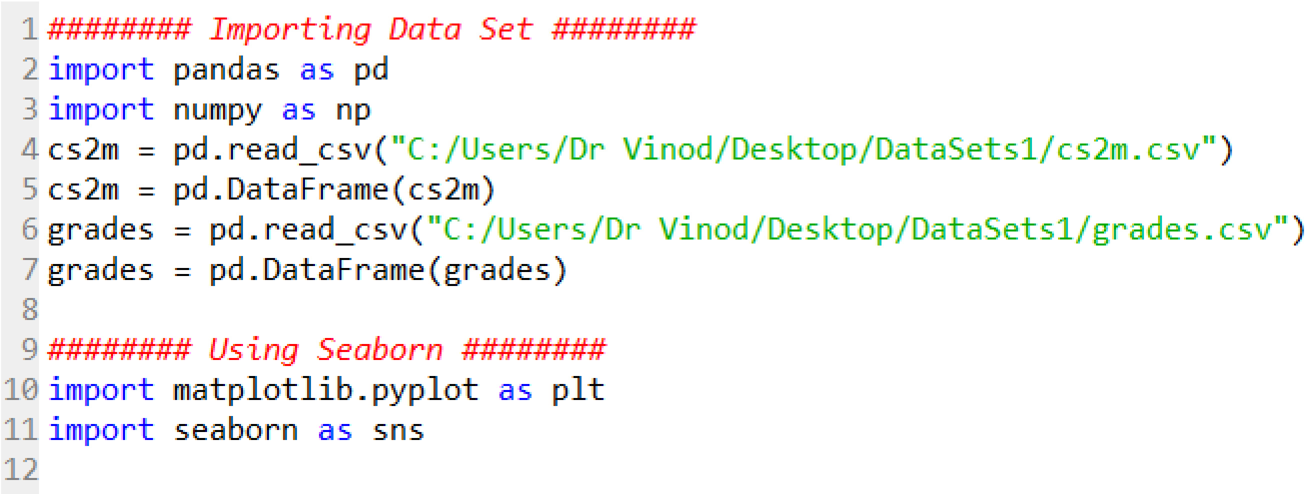
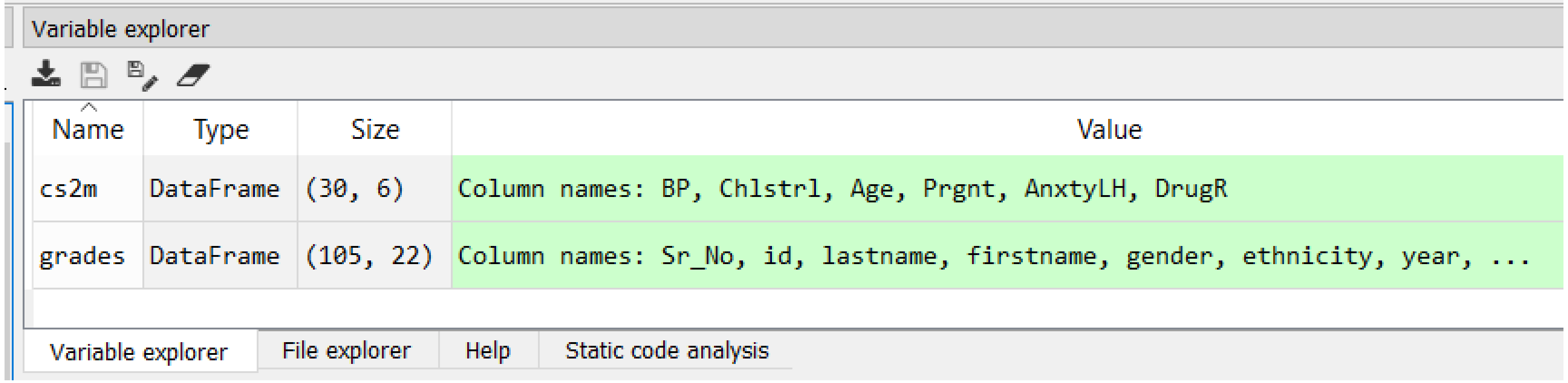
Data Visualization in Python

Data sets: cs2m and grades



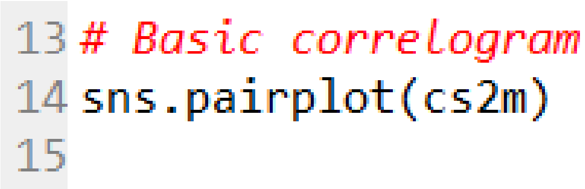
```
######## Importing Data Set ########
import pandas as pd
import numpy as np
cs2m = pd.read_csv("C:/Users/Dr Vinod/Desktop/DataSets1/cs2m.csv")
cs2m = pd.DataFrame(cs2m)
grades = pd.read_csv("C:/Users/Dr Vinod/Desktop/DataSets1/grades.csv")
grades = pd.DataFrame(grades)

######## Using Seaborn ########
import matplotlib.pyplot as plt
import seaborn as sns

```

Import libraries/packages

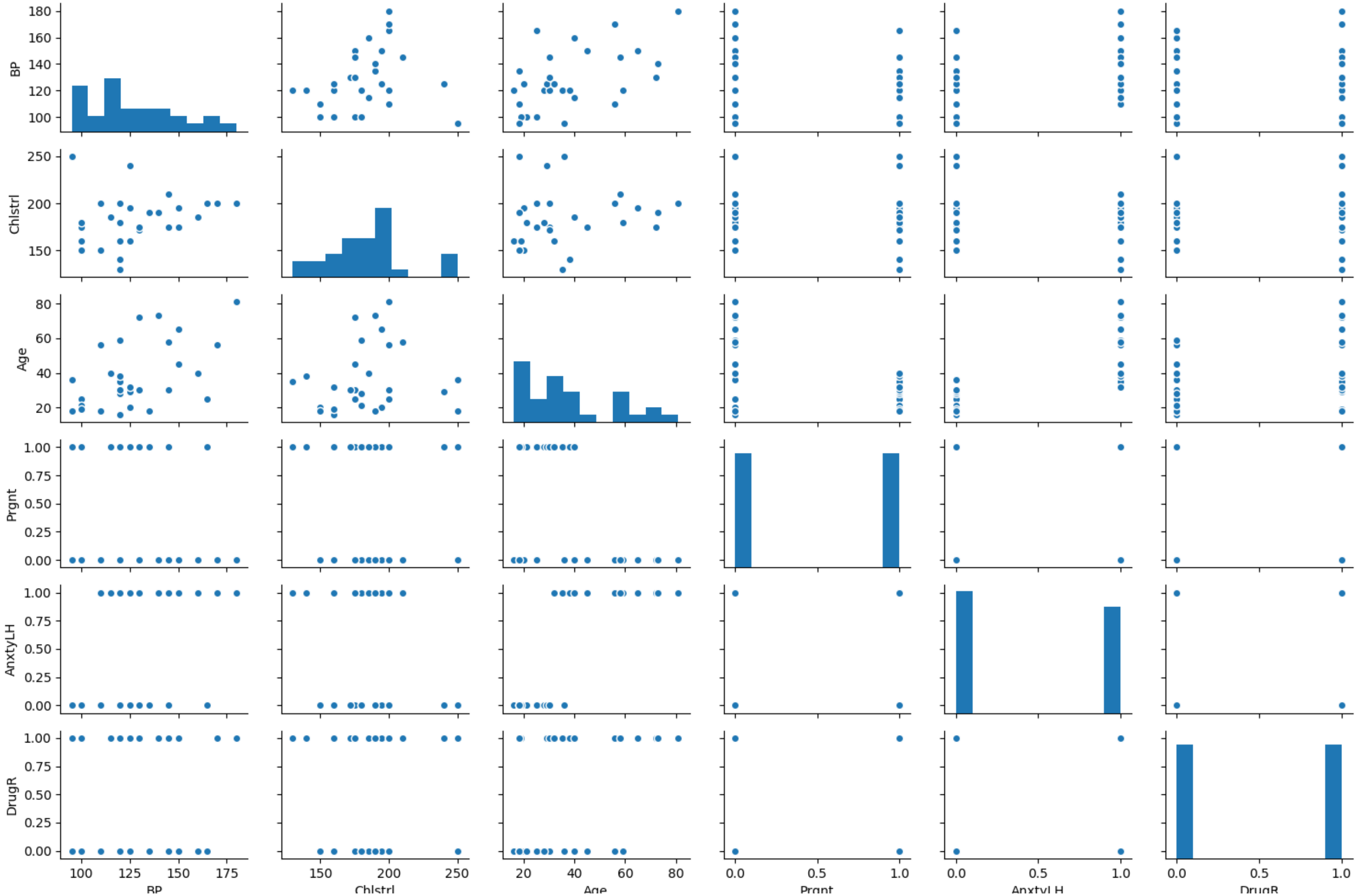


Name	Type	Size	Value
cs2m	DataFrame	(30, 6)	Column names: BP, Chlstrl, Age, Prgnt, AnxtyLH, DrugR
grades	DataFrame	(105, 22)	Column names: Sr_No, id, lastname, firstname, gender, ethnicity, year, ...

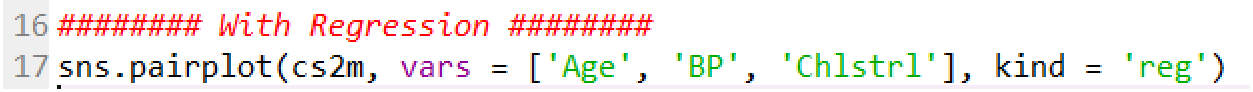
Check files



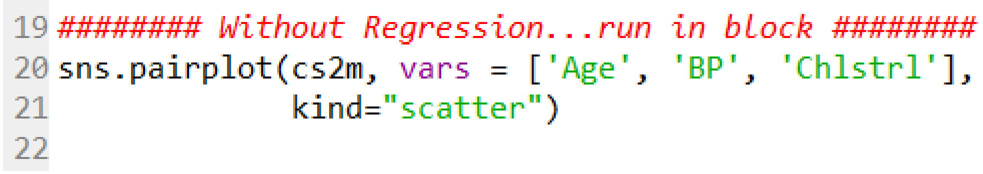
```
# Basic correlogram
sns.pairplot(cs2m)

```



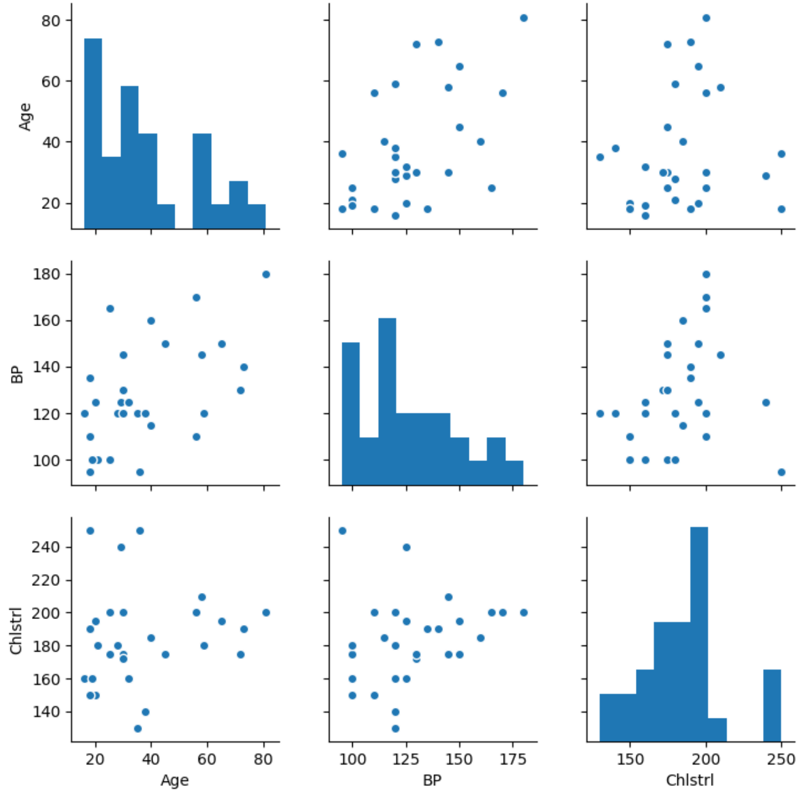
All in one!



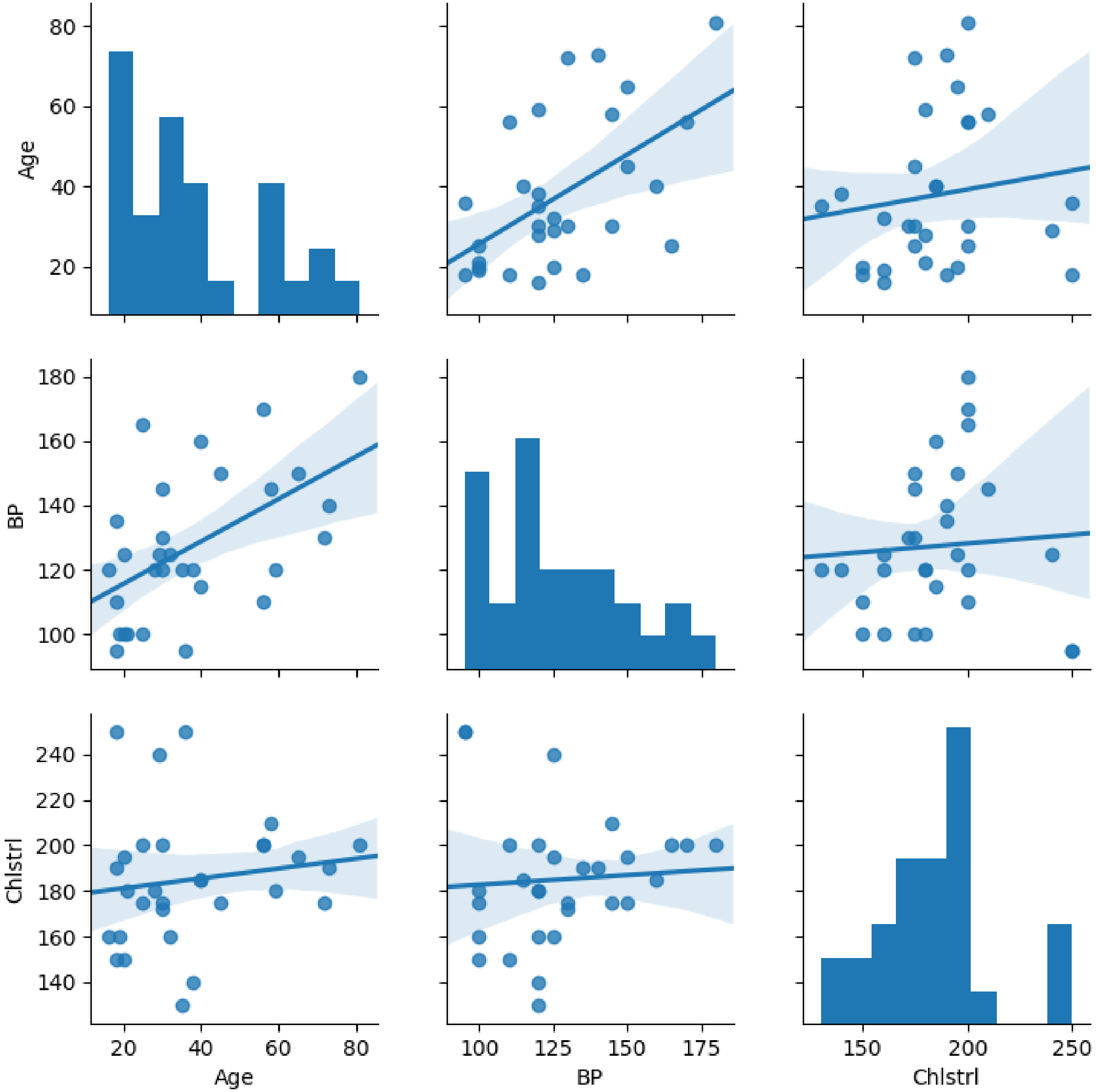
```
######## With Regression ########
sns.pairplot(cs2m, vars = ['Age', 'BP', 'Chlstrl'], kind = 'reg')
```

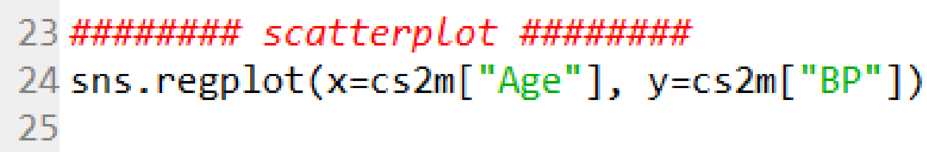


```
######## Without Regression...run in block ########
sns.pairplot(cs2m, vars = ['Age', 'BP', 'Chlstrl'],
                 kind="scatter")
```



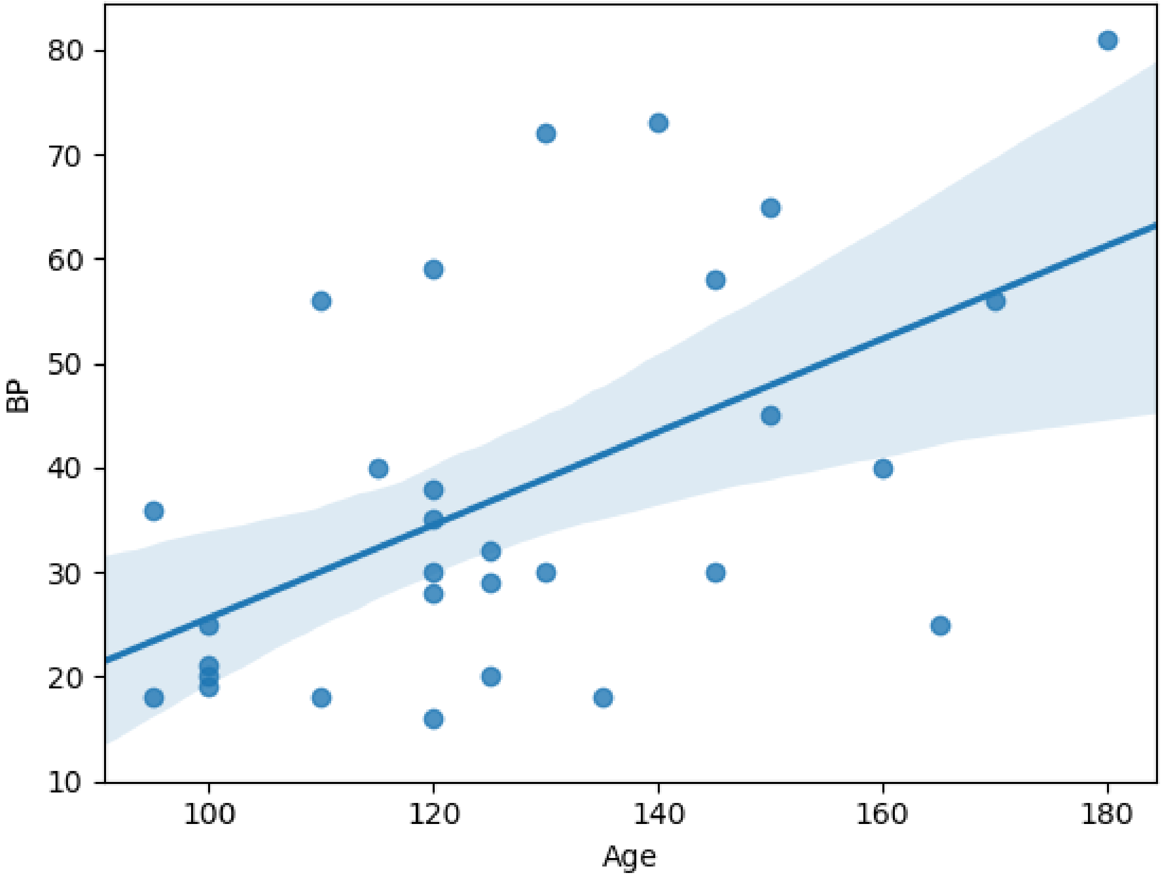
Two continuous variables together

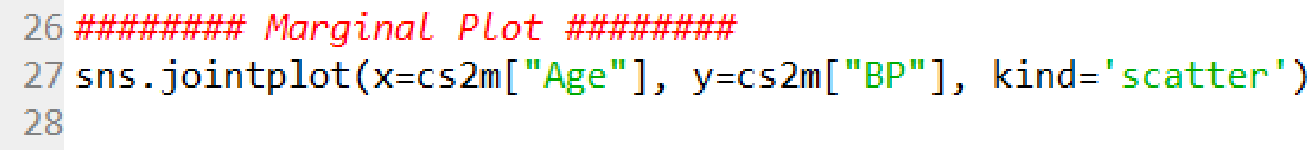




```
######## scatterplot ########
sns.regplot(x=cs2m["Age"], y=cs2m["BP"])

```

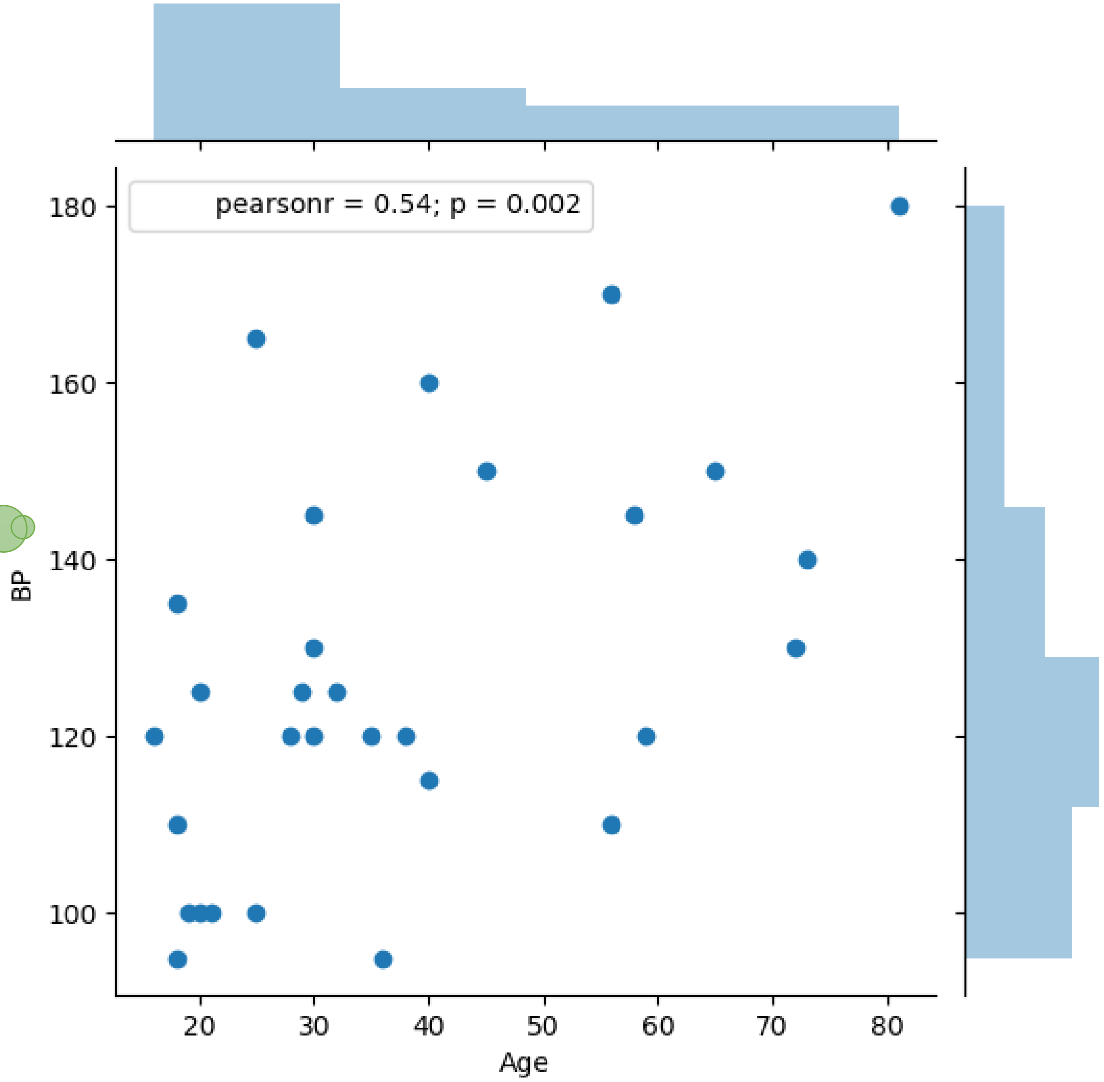
Two continuous variables together



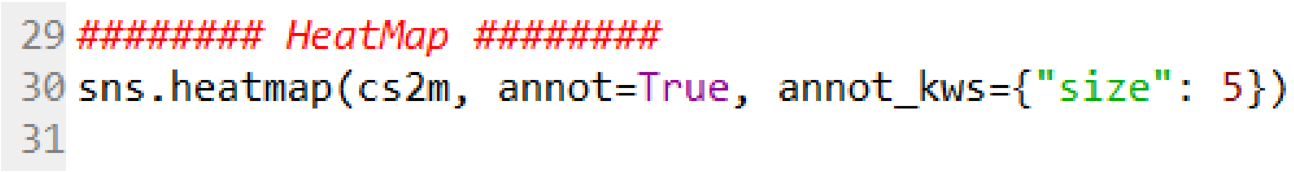


```
######## Marginal Plot ########
sns.jointplot(x=cs2m["Age"], y=cs2m["BP"], kind='scatter')

```

Two continuous variables together

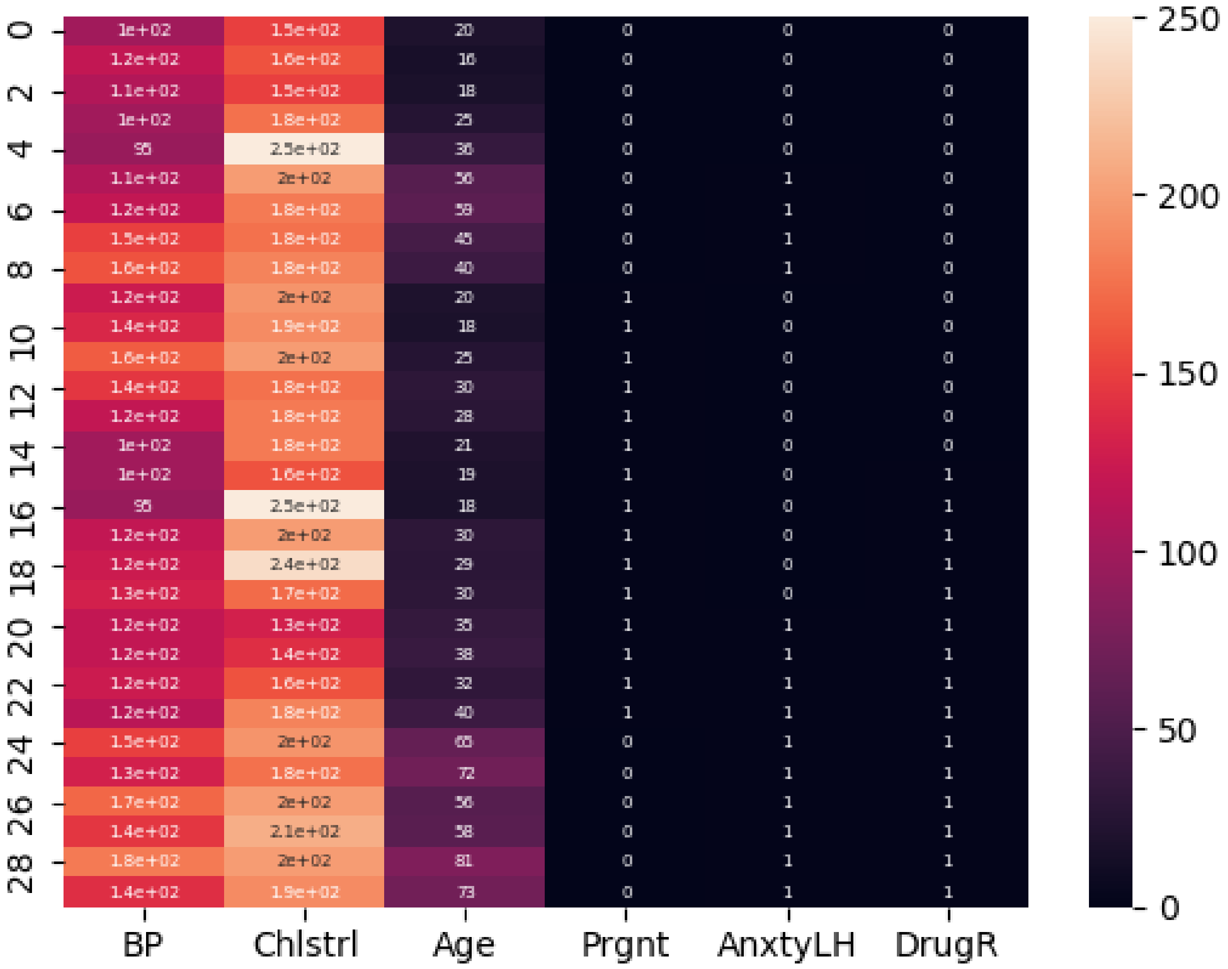


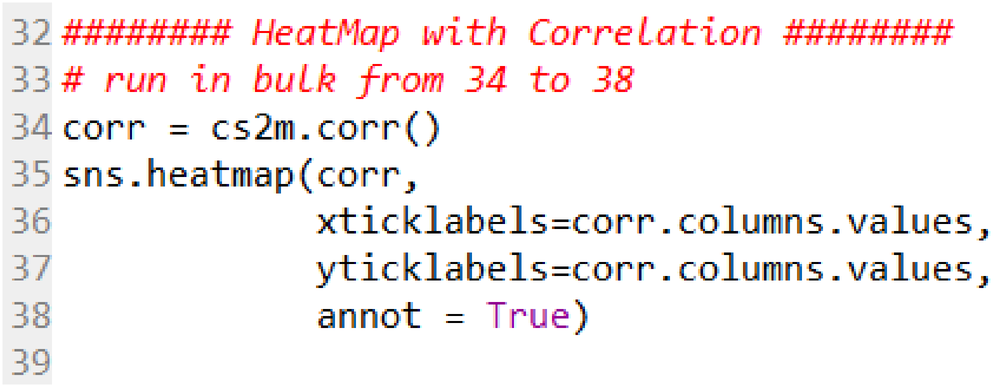
Buy one get one free! Scatter & Histograms!!



```
######## HeatMap ########
sns.heatmap(cs2m, annot=True, annot_kws={"size": 5})

```

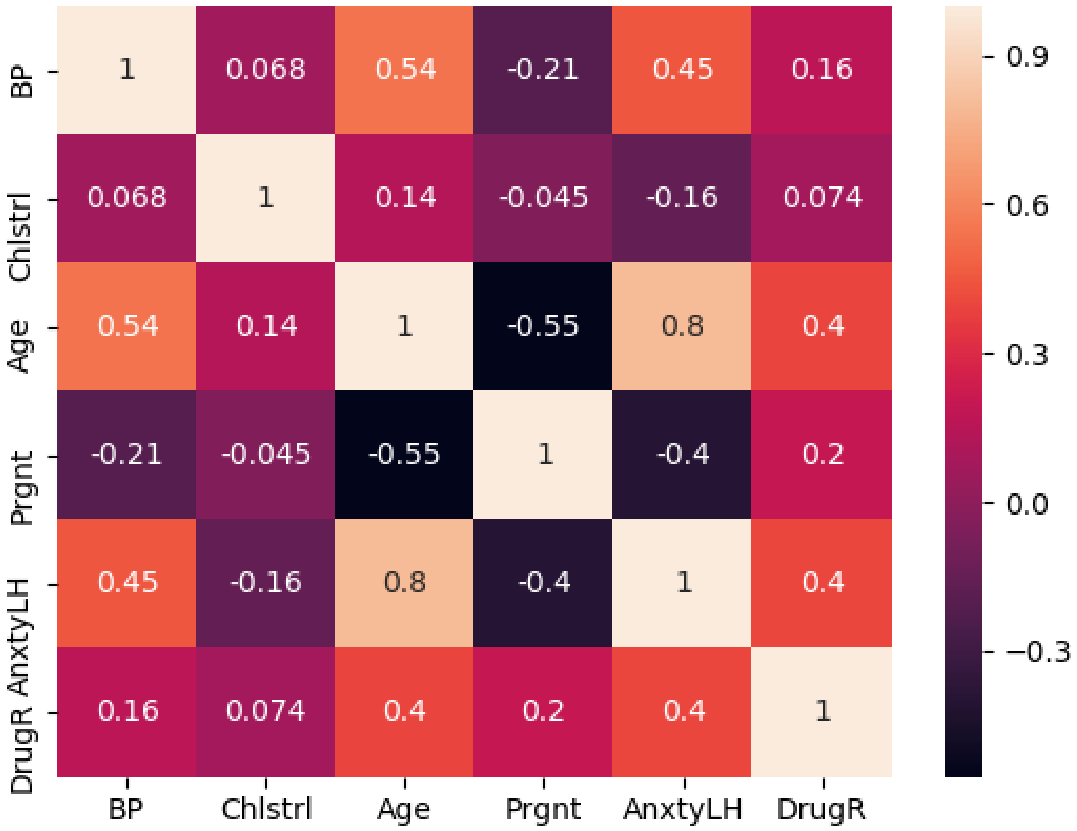
All in one with values 1.3e+02 means 130



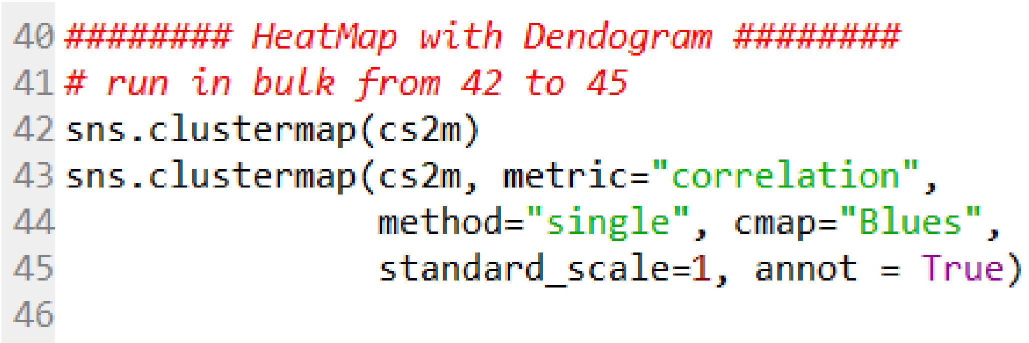


```
32 ######## HeatMap with Correlation ########
33 # run in bulk from 34 to 38
34 corr = cs2m.corr()
35 sns.heatmap(corr,
36             xticklabels=corr.columns.values,
37             yticklabels=corr.columns.values,
38             annot = True)
39
```

Correlation can be between -1 and +1

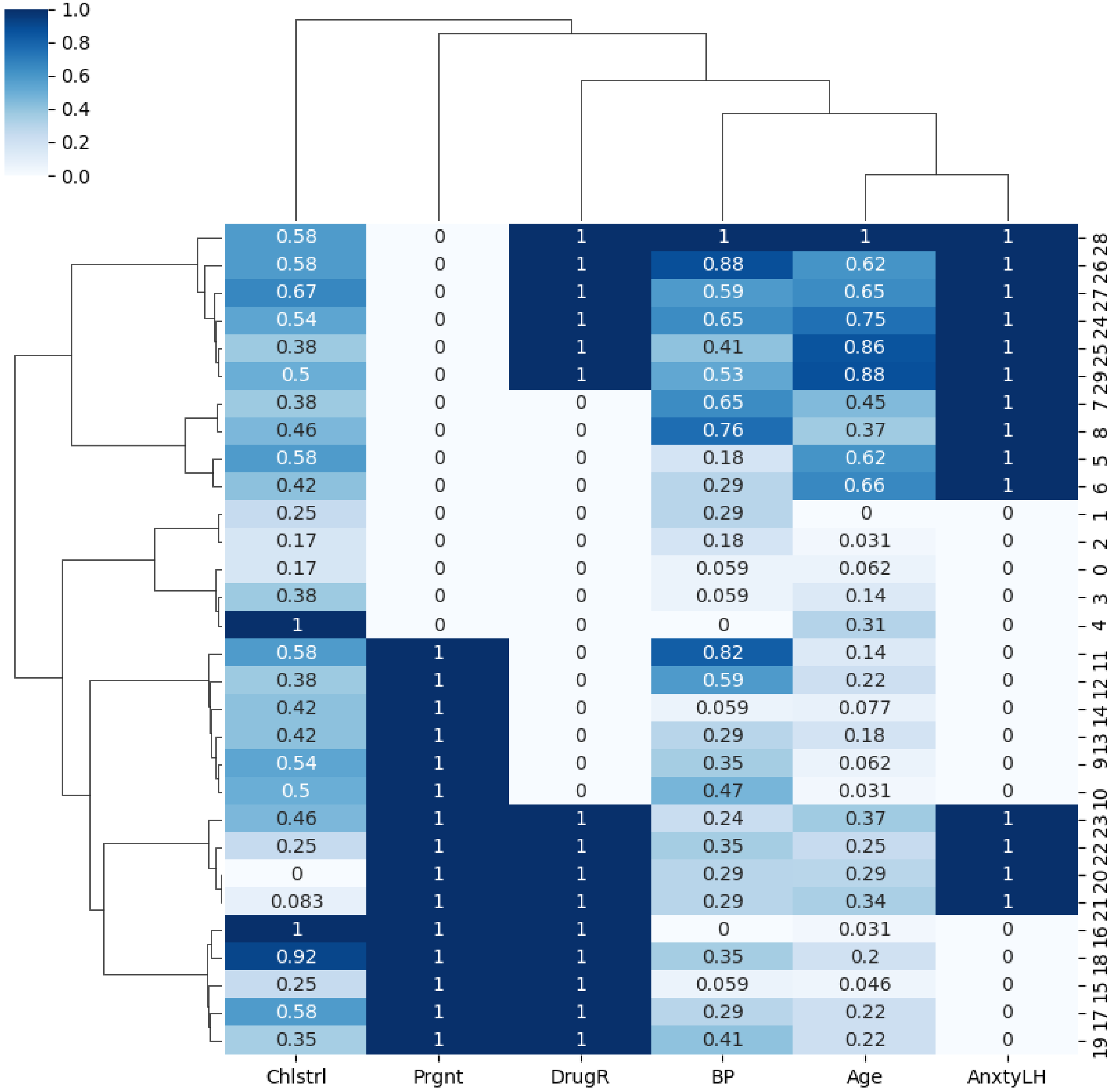


Ignore correlations among categorical variables

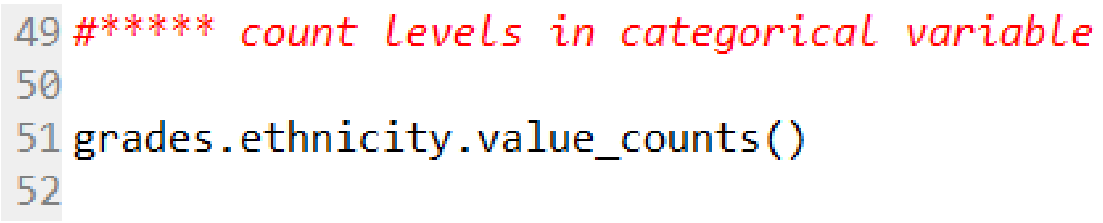


```
40 ######## HeatMap with Dendogram ########
41 # run in bulk from 42 to 45
42 sns.clustermap(cs2m)
43 sns.clustermap(cs2m, metric="correlation",
44                method="single", cmap="Blues",
45                standard_scale=1, annot = True)
46
```

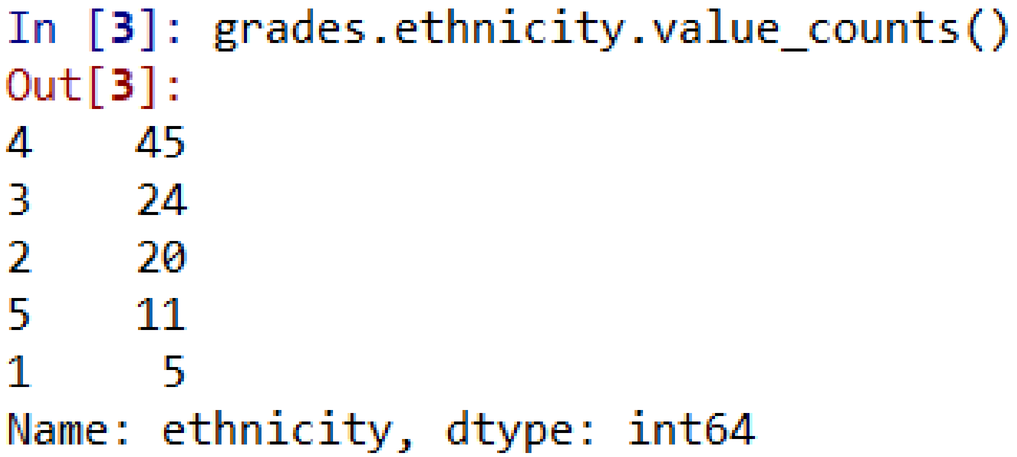
Values are normalized (between 0 and 1)



	Chlstrl	Prgnt	DrugR	BP	Age	AnxtyLH
28	0.58	0	1	1	1	1
26	0.58	0	1	0.88	0.62	1
27	0.67	0	1	0.59	0.65	1
24	0.54	0	1	0.65	0.75	1
25	0.38	0	1	0.41	0.86	1
29	0.5	0	1	0.53	0.88	1
7	0.38	0	0	0.65	0.45	1
8	0.46	0	0	0.76	0.37	1
5	0.58	0	0	0.18	0.62	1
6	0.42	0	0	0.29	0.66	1
1	0.25	0	0	0.29	0	0
2	0.17	0	0	0.18	0.031	0
0	0.17	0	0	0.059	0.062	0
3	0.38	0	0	0.059	0.14	0
4	1	0	0	0	0.31	0
11	0.58	1	0	0.82	0.14	0
12	0.38	1	0	0.59	0.22	0
14	0.42	1	0	0.059	0.077	0
13	0.42	1	0	0.29	0.18	0
9	0.54	1	0	0.35	0.062	0
10	0.5	1	0	0.47	0.031	0
23	0.46	1	1	0.24	0.37	1
22	0.25	1	1	0.35	0.25	1
20	0	1	1	0.29	0.29	1
21	0.083	1	1	0.29	0.34	1
16	1	1	1	0	0.031	0
18	0.92	1	1	0.35	0.2	0
15	0.25	1	1	0.059	0.046	0
17	0.58	1	1	0.29	0.22	0
19	0.35	1	1	0.41	0.22	0

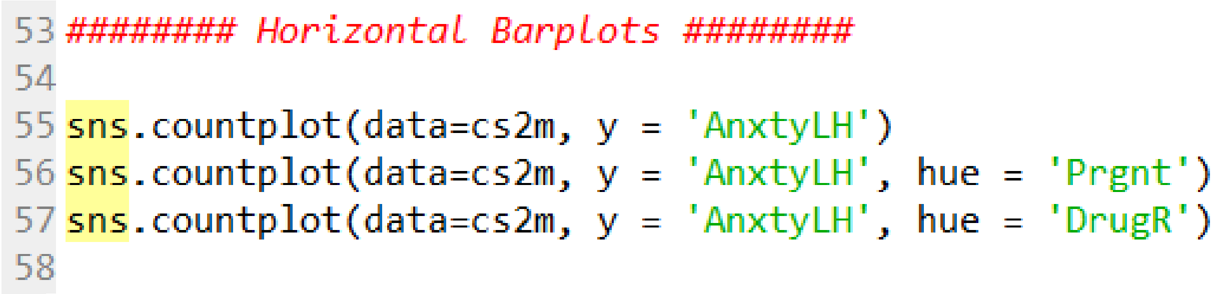


```
#****** count levels in categorical variable

grades.ethnicity.value_counts()

```

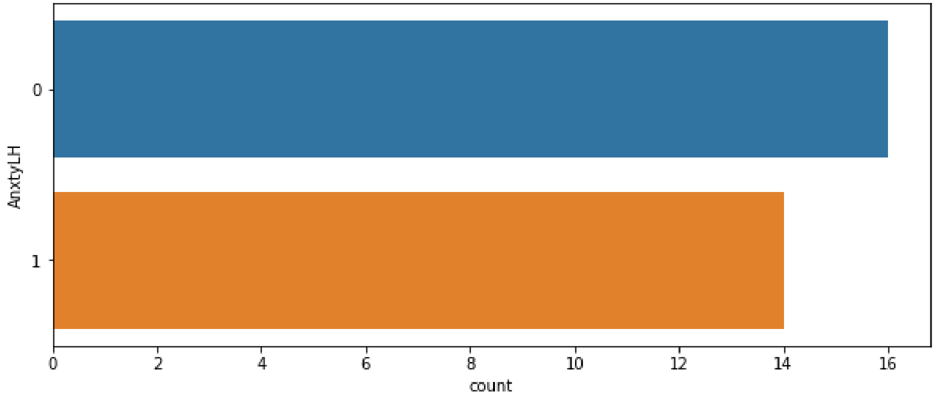


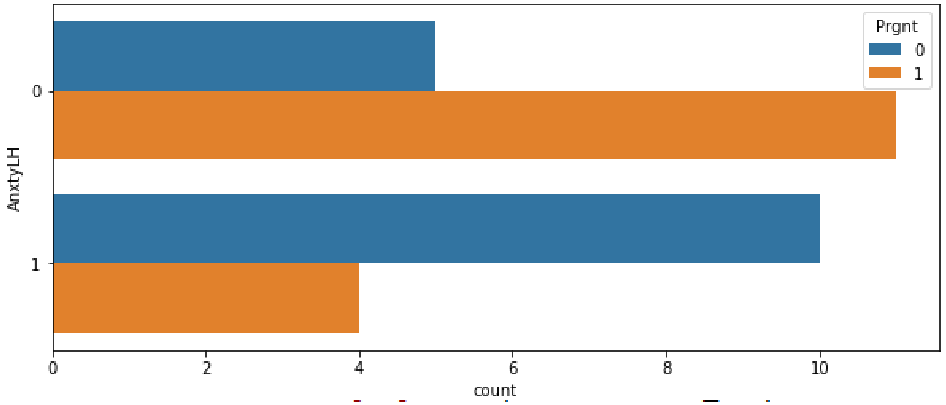
```
In [3]: grades.ethnicity.value_counts()
Out[3]:
4    45
3    24
2    20
5    11
1     5
Name: ethnicity, dtype: int64
```

Counts in each level

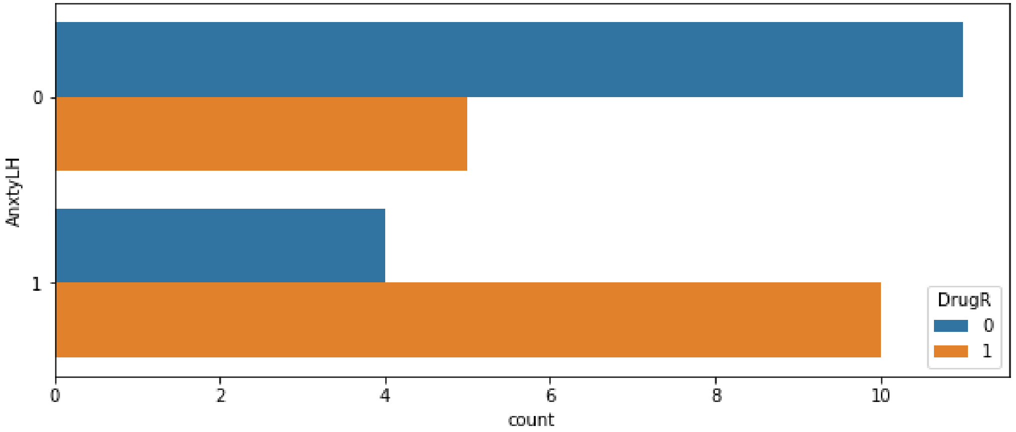


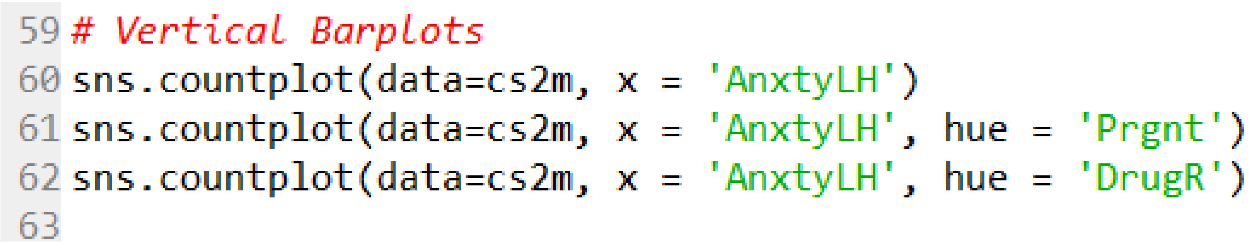
```
######## Horizontal Barplots ########

sns.countplot(data=cs2m, y = 'AnxtyLH')
sns.countplot(data=cs2m, y = 'AnxtyLH', hue = 'Prgnt')
sns.countplot(data=cs2m, y = 'AnxtyLH', hue = 'DrugR')

```



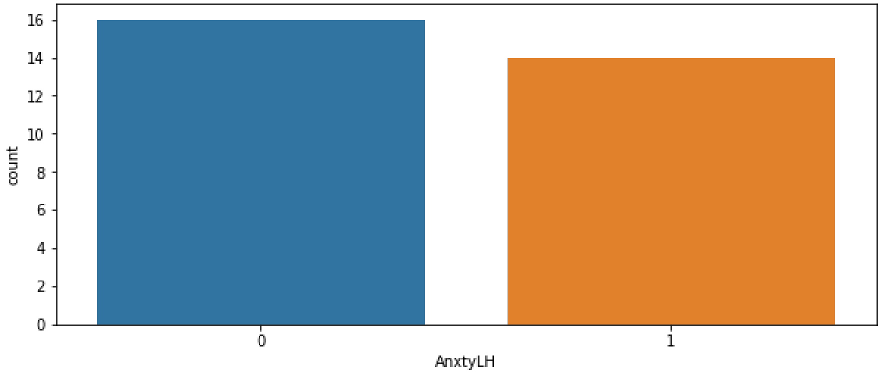


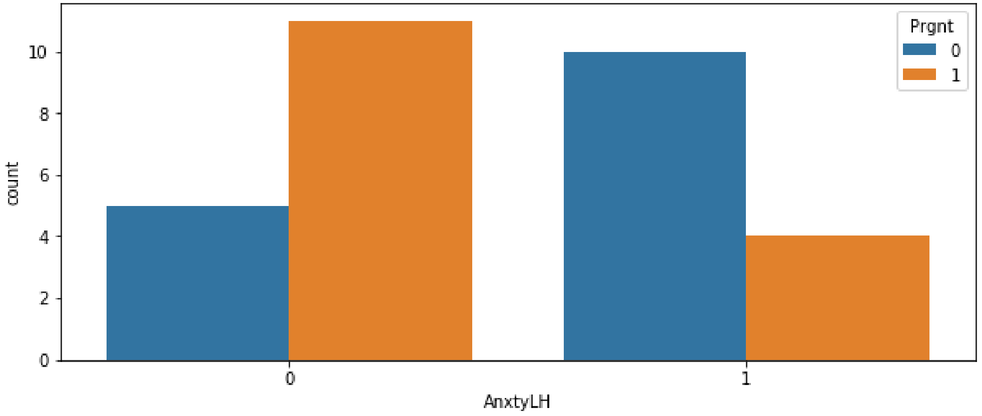
Good idea to see Cross tabulation with these charts



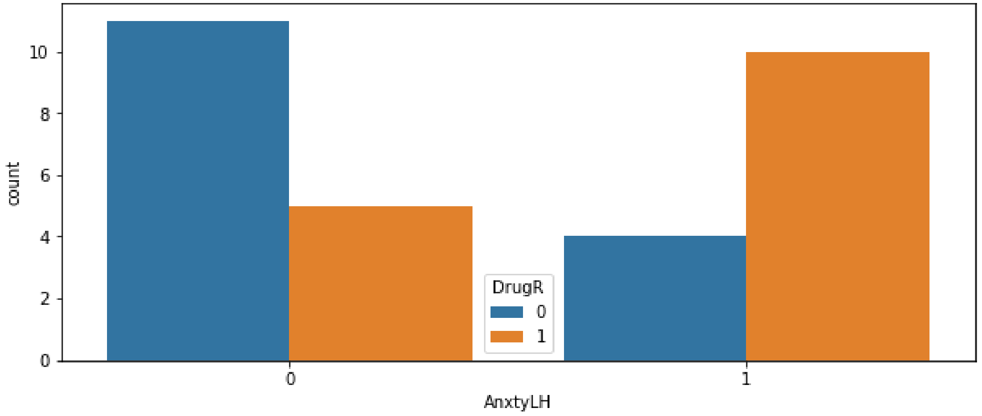


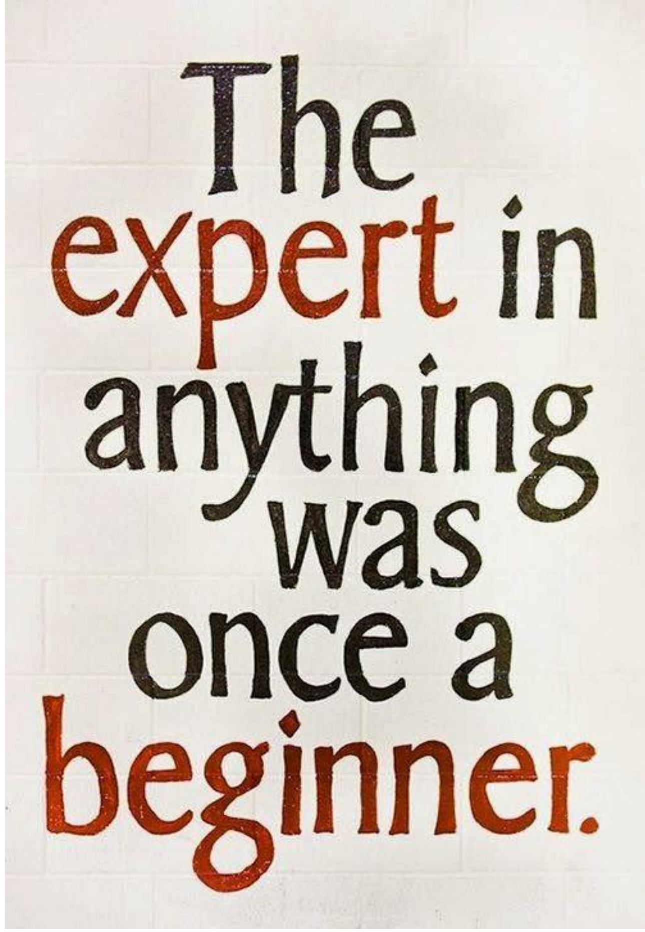
```
# Vertical Barplots
sns.countplot(data=cs2m, x = 'AnxtyLH')
sns.countplot(data=cs2m, x = 'AnxtyLH', hue = 'Prgnt')
sns.countplot(data=cs2m, x = 'AnxtyLH', hue = 'DrugR')

```





Good idea to see Cross tabulation with these charts





The expert in anything was once a beginner.

IT'S NOT ABOUT BEING THE BEST

IT'S ABOUT BEING BETTER THAN YOU WERE YESTERDAY